Effectiveness of Digital Controls in Stopping Academic Dishonesty in Remotely Administered Tests

Emergent Research Forum Paper

Vijay V. Raghavan
Northern Kentucky University
raghavan@nku.edu

Xiaoni Zhang
Northern Kentucky University
zhangx@nku.edu

Abstract

This paper examines the effectiveness of digital controls in mitigating or controlling academic dishonesty in online environments. Conventional methods of administering remote tests range from trust-based assessments to time-controlled tests where there is some control over external help seeking behaviors. Modern learning management systems such as Blackboard also offer tools such as Respondus monitoring though Webcams and LockDown Browsers that prevent browser navigation outside of the test environment. In this study, we compare these two modes of administering tests in a graduate information systems course. The preliminary results suggest that digital monitoring tools are effective controlling help-seeking behaviors.

Keywords

Academic dishonesty, digital controls, distance learning, online learning

Introduction

Universities are offering online courses to supplement their traditional face-to-face offerings to expand their markets. This is invariably accompanied by an increase in assessment procedures conducted remotely. The need to preserve the integrity of tests and test-taking behaviors is paramount in such online environments. Many studies have recognized the differences between online and face-to-face environments, and have examined ways to improve both content delivery and course design in online environments Ryan, Jonick, & Langub, 2015; Aggarwal & Lynn, 2012). Despite an increase in emphasis on assessing for learning outcomes (Watson et al., 2016; Van der Kleij, 2015; Avery, 2003; Bacon, 2003), a corresponding emphasis on testing strategies in online environments is lacking.

Theoretical Foundations

The *theory of planned behavior* links beliefs and behavior (Ajzen, 1985, 2011). The general premise of this theory is that actions are controlled by intentions but not all intentions are carried out; some are discarded while others are adjusted to fit changing conditions. This theory has a strong relevance to our current study. Ethical behavior has been thought to be an evolutionary process, formed by both what we believe and how we behave (Goel et al, 2016). While individual characteristics may predispose a student toward dishonest behaviors during taking tests, the digital controls that are available today have the power of deterring students from a dishonest behavior. Theory of planned behavior has been previously used to predict dishonest behaviors such as tax on-compliance (Bobek at al, 2003). The social cognitive theory of Bandura has provided explanations of human behavior favoring either environmental or internal determinants of behavior (Bandura, 1978; Burnet et al. 2016).

Many studies have explored ways to improve both the *content delivery* and *course design* in online environments (Ryan, Jonick, & Langub, 2015; Aggarwal & Lynn, 2012). Despite an increased emphasis on assessing learning outcomes (Watson et al., 2016; Van der Kleij, 2015; Avery, 2003; Bacon, 2003), a

matching emphasis on testing strategies in online environments is lacking. In a study of about 5000 students in North American universities, it has been found that business students cheat more than their non-business counterparts (McCabe et al, 2006). In situations where in-person proctoring is infeasible, the digital proctoring is the only option available. The question then is how effective are the digital proctoring technologies? This aspect enforcing academic integrity warrants a greater attention as remote administration of tests has become commonplace. Increased online course offerings by academic institutions force us to look for ways to effectively administer tests remotely. Human proctoring of such tests is often impractical and hence digital control of test-taking behavior is necessary. The present study verifies the effectiveness of remote monitoring tools in controlling dishonesty among IS students.

Controlling Cheating During Test-Taking

If tests are given with just enough time to complete, it reduces the possibility of using external aids such as textbooks. In adopting this strategy, there is difficulty in estimating what is "just" enough time and also due to the fact that there might be individual differences in the time estimated. The reliance on digital technologies is critical in remote administration of tests where there are NO peers present during test taking. We consider two such deterrent technologies: *LockDown Browser*® and *Respondus Monitoring*. There are additional test-building features such as randomizing questions from a pool that are available in many Learning Management Systems (LMS) such as Blackboard.

LockDown Browser® is a custom browser that locks down the testing environment. When students use LockDown Browser they cannot navigate to other applications within their computers. They are also prevented from printing, copying, or navigating to another URL. When an assessment is started, students are locked into using this custom browser until they submit it for grading. This is a separate and distinct feature from Respondus monitoring. When using Respondus monitoring the students are being videotaped through their webcam. Students also videotape the environment around them before beginning a test. This ensures that no other helpers either another person or other accessories that aid in performance assessment are present. The teachers can also request the test-takers to show an acceptable ID during this videotaping process to ensure that the person taking the test in a remote location is indeed the one who is enrolled in the course.

H_o1: Subjects in digitally monitored environments (DCE) will have lower test scores than those in the trust-based testing environment (TBE).

Effect of Gender

An interesting question is whether the controls for cheating during tests has the same effect on both genders. Prior studies have shown that women are less likely than men to be dishonest in general when presented with opportunities to cheat and we believe that this behavior can extend to an academic setting as well. The most common explanation for this finding is *sex-role socialization theory* which refers to the process of an individual's behavior, attitudes, and perceptions resemble the societal expectations for persons of his or her gender. According to social learning theory, observational learning also plays a major role in sex-role socialization. By observing the behaviors of others, in real life and in the media, children learn that some behaviors are rewarded in males but not in females and that some behaviors are considered more appropriate for one sex than the other (Bern, 1983; Kohlberg, 1966). Women are socialized to obey the rules, whereas socialization for men is less binding in this respect (Ward & Beck, 1990).

The research question investigated in the current study is whether remote proctoring using digital technologies can control academic dishonesty in test taking? A secondary question whether the students' gender has any effect on altering the predisposition to indulge in help-seeking behaviors. Some previous studies have concluded that cheating behavior is more prevalent in males (Genereux & McLeod, 1995; Sideridis, Tsaousis, & Al Harbi, 2016)

H_o2: The mean scores will be different for Female (F) and male (M) experimental subjects.

There is no reason to believe that digital controls will have different deterrent effect on either gender although we have already hypothesized that help seeking behavior, in general, will vary across genders.

H_o3: Gender and the test-taking environment will not interact to show an effect on test scores.

Experimental Design

One way many teachers counteract possible academic dishonesty is to give time-restricted tests. The presumption here is that if tests are given with just enough time to complete the test, it reduces the possibility of using external aids such as textbooks. This is a simple enough strategy to execute that can be successful but not completely infallible. Fortunately, along with the increased use of digital technologies to administer or take the test there have been improvements in preventive measures that can be initiated digitally. We test two such technologies – Lockdown browser and Respondus monitoring along with test-building features that are available in many online instructional delivery environments such as Blackboard. The experimental and control groups were drawn from the same graduate database course spanning five different semesters. The content covered and test type (True/False) in the two tests was the same. The same instructor taught the class and the instructional design and course delivery methods were the same. The syllabus in each course contained a description of student honor code behavior.

Experimental Task

In the digitally monitored environment, the test questions were selected from a pool that covered the same content areas. The content covered and test type (True/False) in the two tests was the same. There were twenty-five true/false questions drawn from the textbook used for the course. The same instructor covered the class and the instructional design and course delivery methods were the same. The syllabus for each course contained a description of student honor code behavior. In order to mitigate the effect of the negative relationship between the acceptance of rules and actually adhering to them, we provided subsequent reinforcement of expected student behavior the students in the experimental group took the test in a digitally monitored remote environment of their choice. The student in the trust-based environment took the same test using the same digital interface and the only control that was present is that they were given twenty-five questions and the quiz needed to be completed in twenty-five minutes. In the control group, the time to complete was the only deterrent for help-seeking behaviors.

We were testing for the difference in performance between two experimental groups consisting of two different sets of students who received the same test under different testing conditions: one digitally controlled with LockDown Browser and Respondus monitoring and the other purely trust-based with no digital controls. The two conditions were exactly the same except for this difference in the testing environment. Randomly assigning subjects to treatments ensured that all differences between conditions are chance differences.

Experimental Subjects

There were 170 students enrolled in a graduate database class across five different semesters, of which 58 were in the experimental group that used digitally monitored tests and 112 were in the trust-based testing group. 29 in the experimental group were males and 30 were males in the trust-based testing group. The experiment is a two-way unbalanced design. The participants were enrolled in five semesters beginning from Spring 2014 for three years. Through power analysis using G*power, we determined that 134 participants were needed for our design to obtain a high power of 0.95. We randomly selected 134 participants from the total of 170 participants using a random number generator. Thus the participants for data analysis were selected on the basis of random sampling from a naturally occurring pool rather than random assignment to experimental groups. Some participants from this pool were enrolled in a class where the test was administered in a digitally controlled environment. Others took the same test in a trust-based environment where there were no digital controls albeit in different semesters. The content was the same twenty-five T/F questions from a database textbook. Because of the random sampling employed we could not guarantee that we have an equal number of participants in all four conditions constituting the two-way independent sample ANOVA design. Thus, the experiment was a between subjects unbalanced two-way design. The two independent variables were the Gender (male "M" or female "F") and the Group (digitally controlled "D" or trust-based "T") and they both had two levels as noted.

There were 18 females and 24 males in the digitally monitored group and 41 females and 51 males in the trust-based group. We tested for the assumptions of ANOVA by examinng the boxplots to verify that the data is fairly normally distributed and there are no outliers. We ran the Levene Test for the homogeneity of variance and results confirmed that there are no significant differences in homogeneity across conditions ($F(3,112) = 0.4396, p = .7252$).

The question of the type of sum of squares to employ in undertaking an analysis variance of unbalanced data is somewhat controversial (Herr, 1986; Hector, Von Felten, & Schmid, 2010) and the default sum of squares is TYPE III in most packages. The current thinking in conducting statistical analysis of variance favors using TYPE II sum of squares for unbalanced designs as it results in a greater power (Langsrud, 2003; Scholer, 2016; Cooper, 2011).

Cases	Sum of Squares	df	Mean Square	F	p
Gender	52.83	1	32.756	8.1148	0.005
Group	132.68	1	188.997	20.3803	< .001
Gender ✻Group	3.72	1	4.447	0.5712	0.451
Residual	846.32	130	6.771		

Note: Type II Sum of Squares

Table 1: Analysis of Variance Results

TYPE II tests for each main effect after the other main effect. It is based on the assumption that there is no significant interaction (in other words, we should test for interaction first (SS (AB | A, B)) and only if AB is not significant, continue with the analysis for main effects). If there is indeed no interaction, then type II is statistically more powerful than type III. The results show that there is support for the effect that the scores of influenced by gender. Further, a comparison of the means shows that females scored less than the males. This confirms that there is support for females indulged in a lower level of help-seeking behaviors. The direct effect of testing environment is also significant indicating that there is a difference between the digitally monitored and trust-based environments. A comparison of the mean test scores validates that the scores were indeed lower in the digitally monitored environment indicating the digital controls were successful in preventing help-seeking behaviors. However, there is no support for the gender and group interaction. It means that there is no evidence that one gender (either male or female) displayed a differential level of help-seeking behaviors in either environment (digitally monitored or trust-based).

Conclusion

The online education is here to stay. Teaching professionals and students alike have come to appreciate its convenience and flexibility. Need for convenience and flexibility should not lead us to unintentionally making online education less rigorous. In fact, the combination of rigor and convenience seems to strengthen its appeal (Palloff & Pratt, 2008). It is in the hands of teaching professionals to ensure that there is rigor not just in delivering the course content but also in designing and implementing assessment procedures. This requires a good understanding of the technologies available and willingness to implement them. The students who resist changes at the beginning, we believe, will come to appreciate the value that the educators and the institutions are willing to provide for their diplomas. This study has verified how technologies, properly used, can enhance trust in monitored remote testing and consequently in online education itself.

REFERENCES

Aggarwal, A. K., & Lynn, S. A. (2012). Using Continuous Improvement to Enhance an Online Course. *Decision Sciences Journal Of Innovative Education*, *10*(1), 25-48.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From Cognition to Behavior* (pp. 11–39). New York, NY, US: Springer.

Ajzen, I. (2011). The Theory of Planned Behavior: *Handbook of Theories of Social Psychology: Collection: Volumes 1 & 2*, 438.

Avery, E. F. (Ed.). (2003). Assessing student learning outcomes for information literacy instruction in academic institutions. Chicago, IL: Association of College and Research Libraries.

Bandura, A. (1978). The self system in reciprocal determinism. *American psychologist*, 33(4), 344.

Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education*, *25*(1), 31-36.

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-schematic children in a gender-schematic society. Signs, 8(4), 598-616.

Bobek, D. D., & Hatfield, R. C. (2003). An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. *Behavioral Research in Accounting*, 15(1), 13-38.

Buckley, M., Wheeler, A., Mendoza, J., Novicevic, M., Carraher, S., & Beu, D. 2002. Realistic job preview and expectations lowering procedure: A field study. *Journal of Vocational Behavior,* 61: 236–278.

Burnett, A. J., Smith, T. M. E., & Wessel, M. T. (2016). Use of the Social Cognitive Theory to Frame University Students' Perceptions of Cheating. Journal of Academic Ethics, 14(1), 49-69.

Cooper, Matt. (2011, March 2). Anova – Type I/II/III SS explained. Retrieved from https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained/

Hector, A., Von Felten, S., & Schmid, B. (2010). Analysis of variance with unbalanced data: an update for ecology & evolution. *Journal of Animal Ecology*, *79*(2), 308–316.

Herr, D. G. (1986). On the history of ANOVA in unbalanced, factorial designs: The first 30 years. *The American Statistician*, *40*(4), 265-270.

Kohlberg, Lawrence. (1966). A Cognitive-Developmental Analysis of Children's Sex-Role Concepts and Attitudes. In Maccoby, Eleanor E (Ed.), The Development of Sex Differences. Stanford, Calif: Stanford University Press.

Genereux, R. L., & McLeod, B. A. (1995). Circumstances surrounding cheating: A questionnaire study of college students. *Research in Higher Education*, *36*(6), 687-704.

Goel, L., Hart, D., Junglas, I., & Ives, B. (2016). Acceptable {IS} Use: Conceptualization and measurement. *Computers in Human Behavior*, *55, Part A*, 322 – 328.

Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, *13*(2), 163–167.

McCabe, D. L., Butterfield, K. D., & Treviño, L. K. (2006). Academic Dishonesty in Graduate Business Programs: Prevalence, Causes, and Proposed Action. *Academy Of Management Learning & Education*, *5*(3), 294-305.

Palloff, R. M., & Pratt, K. (2008). *Assessing the online learner: Resources and strategies for faculty* (Vol. 14). John Wiley & Sons.

Ryan, M., Jonick, C., & Langub, L. W. (2015). Expert Reflections on Effective Online Instruction: Importance of Course Content. Online Journal of Distance Learning Administration, 18(4).

Scholer, F. (2016, February 5). ANOVA. Retrieved January 5, 2017, from http://goanna.cs.rmit.edu.au/~fscholer/anova.php

Sideridis, G. D., Tsaousis, I., & Al Harbi, K. (2016). Predicting Academic Dishonesty on National Examinations: The Roles of Gender, Previous Performance, Examination Center Change, City Change, and Region Change. *Ethics & Behavior*, *26*(3), 215-237.

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes A Meta-Analysis. Review of educational research, 85(4), 475-511.

Ward, D. A., & Beck, W. L. (1990). Gender and Dishonesty. *The Journal of Social Psychology*, *130*(3), 333–339.

Watson, W. R., Kim, W., & Watson, S. L. (2016). Learning outcomes of a MOOC designed for attitudinal change: A case study of an Animal Behavior and Welfare MOOC. *Computers & Education*, *96*, 83-93.